doi.org/10.54473/IJTRET.2021.5201

International Journal of Trendy Research in Engineering and Technology
Volume 5 Issue 2 April 2021
ISSN No 2582-0958

VIDEO CONTROLLING USING HAND GESTURES FOR DISABLED PEOPLE

Stella Nadar, Simran Nazareth, Kevin Paulson, NilambriNarkar
Department of Computer Engineering, Xavier Institute of Engineering, Mumbai University, Mumbai, India

Received 13 March 2021 Received in revised form 16 March 2021 Accepted 18 March 2021
Available Online 23 March 2021

ABSTRACT

Among many computer visions based interactive systems, designing hand gesture and facial expression based Human-Computer-Interaction (HCI) system retains to be a highly challenging task. Our main purpose is to find a non-tangible way to interact with the computer. Our project aims at developing a video-player controlled by the human hand gestures by making use of Convolutional Neural Network (CNN). The gestures would serve as the direct command for operations such as play or pause the video. It uses simple gestures to control the video. So people don't have to learn the machine like skills and only need to remember a set of gestures to control the video playback.

Keywords—: *Machine Learning, CNN (Convolutional Neural Network), Open CV, Classification model, webcam.*

I. INTRODUCTION

The Hand Gesture recognition is moving at tremendous speed for the futuristic products and services and major companies are developing technology supported the hand gesture system and it includes the devices like Laptop, hand-held devices, Professional and LED lights.The use and adoption will become more cost-effective and cheaper. It's an excellent feature turning data into features with a mixture of technology and Human wave. Smartphones have been experiencing an enormous amount of Gesture Recognition Technology with look and views and working to manage the Smartphone in reading, viewing and that includes what we call touchless gestures. In the medical fields, Hand Gesture can also be experienced in terms of Robotic Nurse and medical assistance. As Technology is usually evolving and changing the longer term is quite unpredictable but we've to be sure the longer term of Gesture Recognition is here to remain with more and eventful and Life touching experiences.

A.Aims and Objectives

There are a lot of critical situations in the day-to-day lives of disabled people. We have come up with a small part of such life to make it easier using computer vision technology. Among many computer visions based interactive systems, designing hand gesture and countenance based HCI system retains to be a highly challenging task. Our main purpose is to find a non-tangible way to interact with the computer. Our project aims at developing a video-player controlled by human hand gestures by making use of Convolutional Neural Network and OpenCV.

Objectives:
1) To minimize the use of a keyboard and mouse in a computer.
2) To integrate gesture recognition features into any computer at a low cost.
3) To help in the development of a non-tangible way to interact with the video player.

B. Scope of the Project

The Hand Gesture recognition is moving at tremendous speed for the futuristic products and services and major companies are developing technology which supports the hand gesture system and it includes the devices like Laptop, hand-held devices. The use and adoption will become less expensive and cheaper. It's an excellent feature turning data into features with a mixture of technology and Human wave. Smartphones have been experiencing an enormous amount of Gesture Recognition Technology with look and views and working to manage the Smartphone in reading, viewing and that includes what we call touchless gestures. In the medical fields, Hand Gesture can also be experienced in terms of Robotic Nurse and medical assistance. As Technology is usually revolving and changing the longer term is quite unpredictable but we've to be sure the longer term of Gesture Recognition is here to remain with more and eventful and Life touching experiences.

C. *Existing System*

Using Convolutional Neural Network (CNN) and Gestures for Human-Computer-Interaction (HCI) may be a pretty recent field of investigation. In that respect, only a couple of articles are published on that topic. The existing system that we are looking at is a video player controlled by hands gesture movements and postures. The development of Human-Computer Interaction (HCI) may be a non-tangible way of communication between a person's and a computer. This project aims at developing a video-player that is controlled by hand gestures.. This application uses a web camera to capture gestures by the user and then perform basic operations based on the input.

II. RELATED WORK

As the aim was to use CNN to achieve the classification, several CNN classification algorithms and related papers were researched for our classification model. Most methods build classifiers based on features computed from the raw inputs. Convolutional neural networks (CNNs) is a deep model that will act directly on the raw inputs. However, such models are currently limited in handling 2D inputs. In this paper, we develop a completely unique 3D CNN model for action recognition. Our model extracts features from both the spatial and the temporal dimensions by performing 3D convolutions, capturing the motion information encoded in multiple adjacent frames.

The developed model generates multiple channels of data from the input frames, and therefore the final feature representation combines information from all channels. To further boost the performance, we propose regularizing the outputs with high-level features and mixing the predictions of a spread of various models.

Understanding an image and classifying it into different hand's gestures using a multiclass classifier, where we predict a class for each image. These images output in the form of an input to a modified VLC system then can allow the system to perform commanded actions. Analysing this gesture is useful in the following ways –

1. For a disabled individual who is unable to walk around in search of the system control devices (remote controller) or to press any keys on the system.

2. For an individual who is unable to understand the upgraded skill of the controller (how to operate the controller).

3. For quick and efficient action on the video running in an emergency.

III. DESCRIPTION

A. *Analysis*

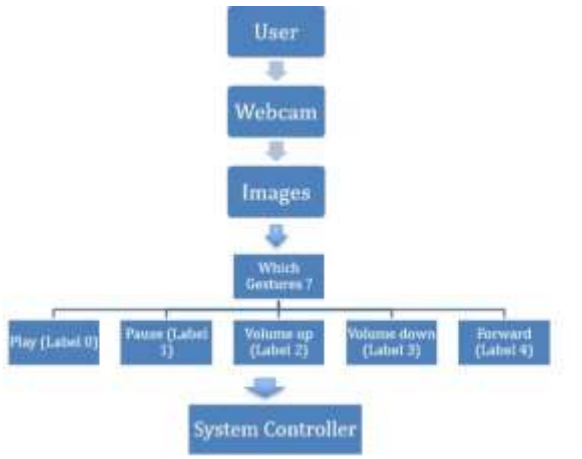

Fig. 1 Basic Architecture of the Project

This is the basic structure that we plan to achieve where the simple human hand gestures will be considered as a control input for the video player to perform actions based on the appropriate mapping of the gestures along with their meaningful control actions.

B. *Feasibility Study*

1) Economic Feasibility: The cost of the project depends upon the training of the model and system requirements for that purpose such as CPU RAM, GPU and disk space. A computer with about 8 GB ram and a basic GPU card was used with shared CPU and GPU processing offered by Google-Colab. The same system was used for the GUI and Android App construction. Hence, by cost-benefit analysis, we can conclude that benefit to cost ratio is high.

2) Technical Feasibility: As mentioned earlier, the task of coding and debugging was made easier by using CNN for the model. As the only webcam to capture images and much GUI is not required, the app construction coding was also not a very difficult task. As every segment will be coded individually and separately, the risk assessment for each segment is easy.

3) Operational Feasibility: The user just has to upload a live video using a webcam. The image just has to be of high quality which can easily be captured from the webcam and processed. Hence, the controls are simple and basic. The efficiency of the project is based upon a few factors, namely, the model and its individual accuracy, the front

end and its latency, the inter-connectivity. All these components can again be controlled individually and hence accuracy of the project as a whole can be easily controlled.

C. Design

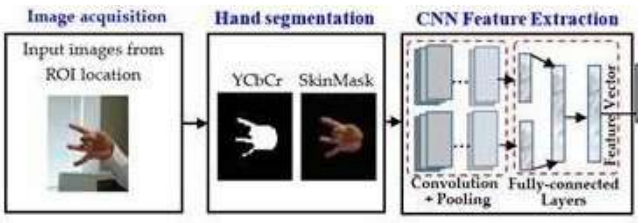

Fig. 2 Flowchart of the Project

Fig. 3 Block Diagram

IV. IMPLEMENTATION

The design for the proposed system is shown in the Figure 3 The images and their masked labels are in .png format. The training images are pre-processed before passing them to the model. The training images and labels are then fed to the CNN model. Tensorflow provides graphs for monitoring the accuracy and loss while training. After training the model it is saved in the HDF5 format. An area is selected for testing the trained model. The prediction is performed on the selected area by loading the trained model which outputs a Class label.

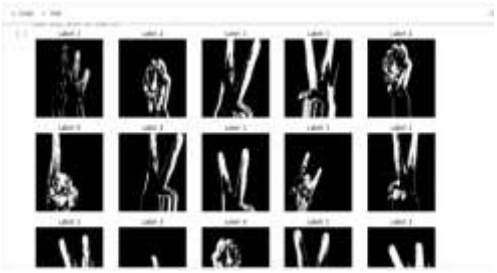

Fig. 4 Gesture Dataset

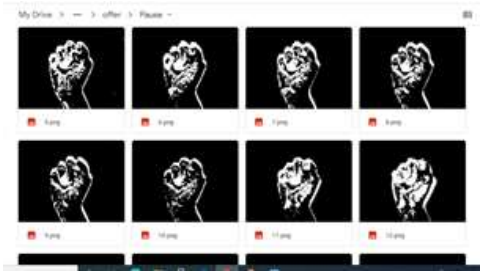

Fig. 5 Pause Gesture

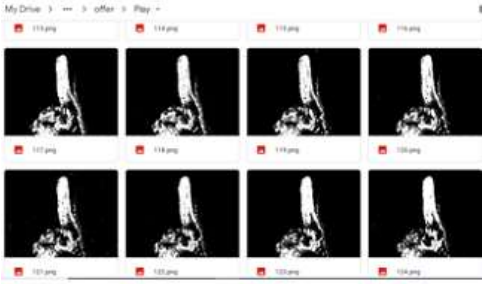

Fig. 6 Play Gesture

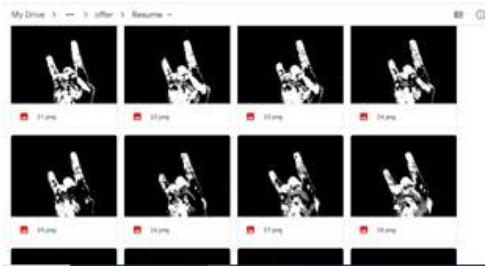

Fig. 7 Resume Gesture

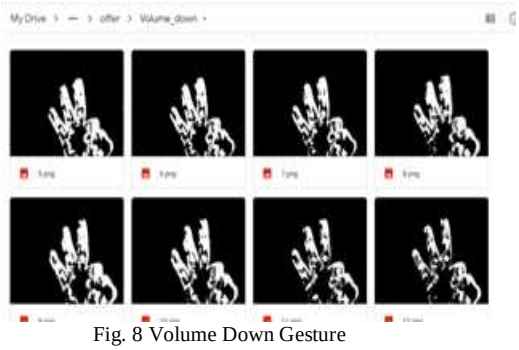

Fig. 8 Volume Down Gesture

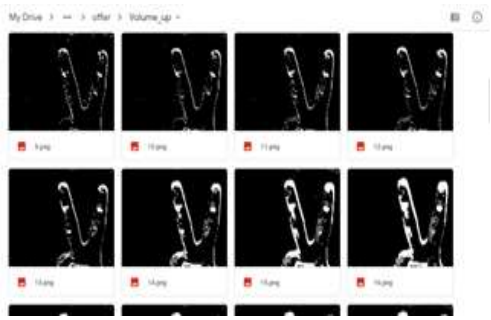

Fig. 9 Volume Up Gesture

After this the gesture pictures were segregated into labelled folders as play, pause, volume up, volume down and resume appropriately. This concluded the dataset acquisition step.

Next, we began working on the classification model. As mentioned earlier CNN or Convolutional Neural Network is a classification algorithm that works best for categorical data/images. It identifies and does the feature extraction work on its own by using the pixel edge detection method, saving us the work to code extraction of individual features. CNN generally consists of 3 layers, Convolution layer, pooling layer and fully connected layer. This structure provides us with a basic blueprint where we only have to choose certain activation functions, values for inbuilt variables, etc. depending upon the desired output.
We used a sequential model with Relu (Rectified Linear Unit) activation. We then converted the 3 channel images (RGB images) into grayscale images with 32 feature detector. In the pooling layer, we are using Max Pooling to reduce the Convolution Matrix. We then add 3 hidden layers with 32, 32 and 64 neurons respectively with Max Pooling. These values and the number of layers were agreed upon after several trial and error run. Next, we use the dropout function to reduce the chances of overfitting. Now we flatten the output matrix and use a dense layer to create a fully connected layer. Sigmoid function along with Softmax activation is used to output values in digit format such as 0, 1, 2, 3, 4 for play, volume up, volume down, resume, pause respectively. Next, we compile the model with Adam optimizer and set the value for epochs and fit the model with train and test data. We trained the model for 50 – 60 epochs. This model gave us an accuracy of about 90.03% against the validation dataset.

Next, we plan to merge this machine learning model with the input stream images captured with a webcam and then the output of this model will be treated as an input to themodified video player.

V. RESULTS

As we see in the following figure, the entire frame captures the hand of the user. On the bottom left of the frame, we can see the output of the gesture.

Fig. 5 Results

VI. PLANS FOR FUTURE

As per our plans, we have completed the back end of the classification model part of the project. Next, we aim to finish the front end or the Video Controller part of the project. A considerable amount of research yet needs to be done in this field as we need to take latency into consideration. We will also work on enhancing the model itself if possible to better perform in real-time. The final aim is to reduce the project response time in real life as even a second delay could cause different control to function.

VII. CONCLUSION

The gesture will serve as the command to perform operations such as play or pause the video based on the user gestures onto the screen.The people only need to remember a set of gestures to control the video playback. The Hand Gesture recognition is moving at tremendous speed for the futuristic products and services and major companies are developing a technology based on the hand gesture system and that includes companies like Microsoft, Samsung, Sony and it includes the devices like Laptop, Handheld devices, Professional and LED lights.

REFERENCES

[1] AnupamAgrawal. "A Vision Based Hand Gesture Interface For Controlling VLC Media Player". In: International Journal of Computer Applications (2010).

[2] Eldose Joy, SruthyChandran, Chikku George, Abhijith A Sabu, DivyaMadhu. "Gesture Controlled Video Player – A non-tangible approach to develop a video player based on Human Hand Gestures using Convolution Neural Networks". In: International Conference in Intelligent Computing and Control Systems (ICICCS) (2018).

[3] SoumikMondalGaurav Sharma. "A Dynamic Hand Gesture Recognition System for Controlling VLC Media Player". In: International Conference on Advances in Technology and Engineering (ICATE). 2013. url: https : / / www . researchgate .net / publication / 261489166_A_dynamic_hand_gesture_recognition_system _for_controlling_VLC_ media_player.

[4] ShuiwangJi, Wei Xu, Ming Yang, Member, Kai Yu. "3D Convolutional Neural Networks for Human Action Recognition". In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2013).